برآورد مدل‌های آمیخته خطی تعمیم‌یافته فضایی با میدان تصادفی چوله گاوسی مانا

فاطمه حسینی[۱]، امید کریمی[۲]

تاریخ دریافت: ۱۴۰۱/۰۵/۰۵
تاریخ پذیرش: ۱۴۰۱/۱۲/۰۲

چکیده:
برای مدل‌بندی داده‌های گسسته فضایی معمولاً از مدل آمیخته خطی تعمیم‌یافته فضایی استفاده می‌شود که همبستگی فضایی در این مدل‌ها با استفاده از متغیرهای پنهان گاوسی وارد مدل می‌شوند. در این مقاله برای افزایش دقت برآورد پارامترها و پیش‌گویی‌ها، متغیرهای پنهان با استفاده از یک میدان تصادفی چوله گاوسی مانا مدل‌بندی و برای برآورد پارامترهای مدل یک الگوریتم جدید بر اساس درستنمایی حاشیه‌ای مرکب ارائه می‌شود. کارایی میدان تصادفی به کار گرفته‌شده و الگوریتم پیشنهادی در یک مثال شبیه‌سازی پیاده‌سازی و بررسی می‌شود.
واژه‌های کلیدی: مدل‌های آمیخته خطی تعمیم‌یافته فضایی، میدان تصادفی چوله گاوسی ایستا، درستنمایی حاشیه‌ای مرکب.

۱ مقدمه

برای تحلیل داده‌های فضایی از نوع گسسته معمولاً از مدل‌های آمیخته خطی تعمیم‌یافته فضایی استفاده و با اضافه کردن متغیرهای پنهان به مدل، وابستگی فضایی موجود در داده‌ها در نظر گرفته می‌شود. یکی از مشکلات اساسی در مدل آمیخته خطی تعمیم‌یافته فضایی این است که به دلیل ناگاوسی بودن متغیرهای پاسخ و وجود متغیرهای پنهان در مدل، تابع درستنمایی شکل بسته‌ای ندارد و به دست آوردن برآورد بیشینه درستنمایی در این مدل‌ها به‌راحتی میسر نیست. همچنین برای سادگی در محاسبات معمولاً متغیرهای پنهان با میدان تصادفی گاوسی مدل‌بندی می‌شوند. در مطالعات فضایی اولین بار [۱۰] میدان تصادفی چوله گاوسی را برای تحلیل داده‌های فضایی چوله معرفی نمود. [۲]، [۱۱] و [۱۲] نیز میدان‌های تصادفی چوله گاوسی بر اساس توزیع چوله نرمال بسته معرفی نمودند. [۱۳] مدل‌های رگرسیون فضایی چوله با داده‌های فضایی گمشده مطالعه نمودند. در مطالعات اشاره‌شده میدان‌های چوله گاوسی معرفی‌شده دارای مشکلاتی مثل عدم مانایی است. یک میدان تصادفی چوله گاوسی تقریباً مانا توسط [۱۴] پیشنهاد شده است که تعمیمی از میدان تصادفی [۲] است. اخیراً [۱] یک میدان تصادفی چوله گاوسی مانا معرفی نمودند. برای اولین بار [۴] استفاده از میدان تصادفی چوله گاوسی را برای متغیرهای پنهان در مدل‌های آمیخته خطی تعمیم‌یافته فضایی پیشنهاد کردند. سپس [۵]، [۸]، [۶]، [۷] و [۹] در مدل‌های آمیخته خطی تعمیم‌یافته فضایی با متغیرهای پنهان فضایی چوله، با معرفی الگوریتم‌های کارای مختلف و مقایسه آن‌ها به برآورد این مدل پرداختند. [۹] یک تحلیل درستنمایی مرکب ([۱۵])، برای داده‌های فضایی با متغیرهای پنهان چوله بیان کردند که از میدان تصادفی چوله گاوسی تقریباً مانا [۱۴] استفاده کرده‌اند.

در این مقاله از یک میدان تصادفی چوله گاوسی مانا برای مدل‌بندی متغیرهای پنهان فضایی استفاده می‌شود و پس از معرفی مدل رهیافت درستنمایی معرفی‌شده در [۹] به این مدل تعمیم داده می‌شود. ساختار مقاله به این صورت است که در بخش دوم تعریف‌هایی از توزیع چوله نرمال بسته و میدان تصادفی چوله گاوسی تقریباً مانا ارائه می‌شوند. در بخش سوم میدان تصادفی چوله گاوسی مانا تعریف می‌شود. مدل آمیخته خطی تعمیم‌یافته فضایی بر اساس میدان تصادفی چوله گاوسی مانا در بخش چهارم معرفی و برآورد درستنمایی مدل بیان می‌شود. در نهایت در بخش پنجم کارایی مدل معرفی‌شده با میدان تصادفی چوله گاوسی مانا و میدان تصادفی گاوسی و رهیافت درستنمایی ارائه‌شده در یک مثال شبیه‌سازی موردبحث و بررسی قرار می‌گیرد.

[۱] دانشگاه سمنان، گروه آمار. (نویسنده مسئول: fatemeh.hoseini@semnan.ac.ir)
[۲] دانشگاه سمنان، گروه آمار

۲ میدان تصادفی چوله گاوسی بسته تقریباً مانا

بردار تصادفی n بعدی $\boldsymbol{X}$ دارای توزیع چوله نرمال بسته چند متغیره با پارامترهای $\boldsymbol{\mu}$، $\boldsymbol{\Sigma}$، $\boldsymbol{\Gamma}$، $\boldsymbol{\nu}$ و $\boldsymbol{\Delta}$ است اگر تابع چگالی آن به‌صورت

$$\begin{aligned} f_{n,q}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Gamma},\boldsymbol{\nu},,\boldsymbol{\Delta}) &= k\phi_n(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) \\ &\times \Phi_q(\boldsymbol{\Gamma}(\boldsymbol{x}-\boldsymbol{\mu});\boldsymbol{\nu},\boldsymbol{\Delta}), \quad (۱) \end{aligned}$$

باشد که در آن $k=[\Phi_q(\circ;\boldsymbol{\nu},\boldsymbol{\Delta}+\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top)]^{-1}$، $\boldsymbol{\mu}$ بردار پارامتر مکان، ماتریس $n\times n$ معین مثبت $\boldsymbol{\Sigma}$ ماتریس مقیاس، عناصر ماتریس $\boldsymbol{\Gamma}_{q\times n}$ پارامترهای چولگی هستند. $\Phi_q(\boldsymbol{\Gamma}(\boldsymbol{x}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta)$ تابع توزیع تجمعی q متغیره نرمال با بردار میانگین $\boldsymbol{\nu}$ و ماتریس واریانس کوواریانس $\boldsymbol{\Delta}$ است. به‌طور خلاصه این توزیع به‌صورت $CSN_{n,q}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Gamma},\boldsymbol{\nu},\boldsymbol{\Delta})$، نمایش داده می‌شود. گشتاور مرتبه‌ی اول توزیع چوله نرمال بسته به‌صورت

$$\mathrm{E}(\boldsymbol{X})=\boldsymbol{\mu}+\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top\boldsymbol{\Psi} \quad (۲)$$

به دست می‌آید که در آن

$$\begin{aligned} \boldsymbol{\Psi} &= \boldsymbol{\Psi}(\circ,\boldsymbol{\Sigma},\boldsymbol{\Gamma},\boldsymbol{\nu},\boldsymbol{\Delta}) \\ &= \frac{\Phi_q^*(\circ;\boldsymbol{\nu},\boldsymbol{\Delta}+\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top)}{\Phi_q(\circ;\boldsymbol{\nu},\boldsymbol{\Delta}+\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top)} \quad (۳) \end{aligned}$$

است. برای ماتریس معین مثبت $\boldsymbol{\Omega}$،

$$\Phi_q^*(\boldsymbol{s};\boldsymbol{\nu},\boldsymbol{\Omega})=[\nabla_s\Phi_q(\boldsymbol{s};\boldsymbol{\nu},\boldsymbol{\Omega})]^\top,$$

که در آن

$$\nabla_s=(\frac{\partial}{\partial s_۱},\cdots,\frac{\partial}{\partial s_q})^\top$$

می‌باشد. واریانس توزیع چوله نرمال بسته به‌صورت

$$\begin{aligned} \mathrm{Var}(\boldsymbol{X}) &= \mathrm{E}(\boldsymbol{X}\boldsymbol{X}^\top)-(\mathrm{E}(\boldsymbol{X}))(\mathrm{E}(\boldsymbol{X}))^\top, \\ &= \boldsymbol{\Sigma}+\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Sigma}-\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top\boldsymbol{\Psi}\boldsymbol{\Psi}^\top\boldsymbol{\Gamma}\boldsymbol{\Sigma} \quad (۴) \end{aligned}$$

است که در آن

$$\begin{aligned} \boldsymbol{\Lambda} &= \boldsymbol{\Lambda}(\circ,\boldsymbol{\Sigma},\boldsymbol{\Gamma},\boldsymbol{\nu},\boldsymbol{\Delta}) \\ &= \frac{\Phi_q^{**}(\circ;\boldsymbol{\nu},\boldsymbol{\Delta}+\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top)}{\Phi_q(\circ;\boldsymbol{\nu},\boldsymbol{\Delta}+\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^\top)} \quad (۵) \end{aligned}$$

و $\Phi_q^{**}(\boldsymbol{s};\boldsymbol{\nu},\boldsymbol{\Omega})=\nabla_s\nabla_s^\top\Phi_q(\boldsymbol{s};\boldsymbol{\nu},\boldsymbol{\Omega})$ تعریف می‌شوند.

میدان تصادفی چوله گاوسی بسته توسط [۱۴] به این صورت تعریف شد که فرض کنید $\boldsymbol{U}(\boldsymbol{s})=\{(U_۱(\boldsymbol{s}),U_۲(\boldsymbol{s}))^\top,\boldsymbol{s}\in\mathcal{D}\}$ میدان تصادفی گاوسی دومتغیره باشد که در آن D یک مجموعه از اعداد حقیقی به‌صورت $D\subseteq R^d$ و $\boldsymbol{U}_۲=[U_۲(\boldsymbol{s}'_۱),\cdots,U_۲(\boldsymbol{s}'_q)]$ آنگاه میدان تصادفی چوله گاوسی بسته به شکل

$$\{\boldsymbol{X}(\boldsymbol{s})=[U_۱(\boldsymbol{s})|\boldsymbol{U}_۲\leq\boldsymbol{0}\},$$

تعریف می‌شود، اگر برای هر مجموعه متناهی $(\boldsymbol{s}_۱,\cdots,\boldsymbol{s}_n)$،

$$\boldsymbol{X}=(X(\boldsymbol{s}_۱),\cdots,X(\boldsymbol{s}_n))^\top,$$

دارای توزیع چوله نرمال بسته باشد. در این میدان تصادفی اگر موقعیت‌های $(\boldsymbol{s}'_۱,\cdots,\boldsymbol{s}'_q)$ به‌اندازه‌ی کافی از مرزها دور، $n=q$ و $(\boldsymbol{s}'_۱,\cdots,\boldsymbol{s}'_n)=(\boldsymbol{s}_۱,\cdots,\boldsymbol{s}_n)$ باشند و توزیع چوله نرمال بسته برای $\boldsymbol{X}$ به شکل

$$CSN_{n,n}(\boldsymbol{\mu},\sigma^۲\boldsymbol{C}_\varphi,\frac{\gamma}{\sigma}\mathbf{I_n},\nu\mathbf{1_n},(۱-\gamma^۲)\mathbf{I_n}), \quad (۶)$$

در نظر گرفته شود، آنگاه $\boldsymbol{X}(\boldsymbol{s})$ تقریباً مانا است چون این میدان تصادفی همانند میدان تصادفی گاوسی دارای خاصیت حاشیه‌ای نیست و توزیع‌های حاشیه‌ای آن برای هر تحقق p تایی به کل ناحیه فضایی وابسته است. در (۶)، $\boldsymbol{\mu}$ پارامتر مکان، $\sigma^۲$ پارامتر مقیاس، $|\gamma|<۱$ پارامتر چولگی، $\boldsymbol{C}_\varphi$ ماتریس همبستگی مانا، $\mathbf{I_n}$ ماتریس همانی و $\mathbf{1_n}\in\mathbb{R}^\mathbf{n}$ برداری با عناصر یک است. چگالی حاشیه‌ای X_j به ماتریس همبستگی $\boldsymbol{C}_\varphi$ که مربوط به کل شبکه فضایی $(\boldsymbol{s}_۱,\cdots,\boldsymbol{s}_n)$ است وابسته است، اما ساختار کلی این چگالی یک توزیع چوله نرمال بسته است. ازاین‌رو [۱۴] آن را یک میدان تصادفی تقریباً مانا نامیدند، برای مطالعه بیش‌تر به [۹] و [۱۴] مراجعه شود.

۳ میدان تصادفی چوله گاوسی بسته مانا

در قسمت قبل دیدیم میدان تصادفی چوله گاوسی که توسط [۱۴] معرفی‌شده است تقریباً مانا است، به‌عنوان‌مثال [۱] نشان داد توزیع حاشیه‌ای X_j به ماتریس همبستگی $\boldsymbol{C}_\varphi$ که مربوط به کل شبکه فضایی $(\boldsymbol{s}_۱,\cdots,\boldsymbol{s}_n)$ می‌باشد وابسته است، اما ساختار کلی چگالی X_j یک توزیع چوله نرمال بسته است. برای رفع این مشکل یعنی برقراری شرط سازگاری حاشیه‌ای و تعریف یک میدان تصادفی چوله گاوسی مانا، [۱] یک تحقق nتایی از میدان تصادفی چوله گاوسی تقریباً مانا به‌صورت $\boldsymbol{X}=(X(\boldsymbol{s}_۱),\ldots,X(\boldsymbol{s}_n))^\top$ مطابق رابطه (۶) با تفکیک‌های زیر

$$\boldsymbol{X}_{n\times ۱}=\begin{pmatrix}\boldsymbol{X}_۱\\ \boldsymbol{X}_۲\end{pmatrix},\boldsymbol{C}_\varphi=\begin{pmatrix}C_{۱۱} & C_{۱۲}\\ C_{۲۱} & C_{۲۲}\end{pmatrix},$$

در نظر گرفتند که در آن $\boldsymbol{X}_1 \in R^{n_1}$، $\boldsymbol{X}_2 \in R^{n_2}$ و $n_1+n_2=n$ است و ماتریس همبستگی فضایی $\boldsymbol{C}_\varphi$ تفکیک‌های متناظر با بردارهای $\boldsymbol{X}_1$ و $\boldsymbol{X}_2$ از ماتریس $\boldsymbol{X}$ هستند. سپس با به دست آوردن تابع مولد گشتاور حاشیه‌ای $\boldsymbol{X}_1$ نشان دادند که تابع مولد گشتاور حاشیه‌ای $\boldsymbol{X}_1$ به ماتریس همبستگی $\boldsymbol{C}_\varphi$ برای کل ناحیه فضایی وابسته است و شرط سازگاری حاشیه‌ای برقرار نیست. برای این‌که شرط سازگاری حاشیه‌ای برقرار شود، [۱] پیشنهاد نمودند که پارامترهای یک تحقق از میدان تصادفی چوله گاوسی تقریباً مانا را به‌صورت

$$\boldsymbol{X} \sim CSN_{n,n}\big(\mu \mathbf{1}_n, \sigma^2 \boldsymbol{C}_\varphi, \frac{\gamma}{\sigma}\boldsymbol{C}_\varphi{}^{-\frac{1}{2}}, \nu \mathbf{1}_n, (1-\gamma^2)\mathbf{I}_n\big), \quad (۷)$$

اصلاح شوند که در آن $\boldsymbol{C}_\varphi{}^{-\frac{1}{2}}$ عکس ریشه دوم ماتریس $\boldsymbol{C}_\varphi$ است. با محاسبه تابع مولد گشتاور توزیع حاشیه‌ای $\boldsymbol{X}_1$ نشان دادند که توزیع حاشیه‌ای $\boldsymbol{X}_1$ برای میدان تصادفی چوله گاوسی تعریف‌شده در رابطه (۷) دارای بعد n_1 و از خانواده توزیع‌های چوله نرمال بسته است، پس دارای شرط سازگاری حاشیه‌ای و یک میدان تصادفی خوش‌تعریف می‌باشد.

۴ مدل

فرض کنید $\boldsymbol{X}=(x_1,\cdots,x_n)^\top$ یک تحقق از میدان تصادفی گاوسی چوله بسته در n موقعیت $\{s_1,\cdots,s_n\}$ با دامنه $\mathcal{D}\subseteq \mathbb{R}^n$ و دارای توزیع چوله نرمال به‌صورت (۷) با پارامتر مکان $\boldsymbol{\mu}=\boldsymbol{H\beta}$ باشد که در آن $\boldsymbol{H}$ ماتریسی با بعد $n\times(p+1)$ شامل متغیرهای کمکی و $\boldsymbol{\beta}$ بردار پارامترهای رگرسیونی است. برای ماتریس $\boldsymbol{C}_\varphi$ ساختار فضایی نمایی همسانگرد فرض می‌شود. از رابطه‌ی (۱) تابع چگالی $\boldsymbol{X}$ به‌صورت

$$\begin{aligned} f(\boldsymbol{x}|\boldsymbol{\eta}) &= \phi_n(\boldsymbol{x};\boldsymbol{H\beta},\sigma^2\boldsymbol{C}_\varphi) \\ &\times \frac{\Phi_n(\frac{\gamma}{\sigma}C_\varphi{}^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{H\beta}),\nu\mathbf{1}_n,(1-\gamma^2)\mathbf{I}_n)}{\Phi_n(\mathbf{0};\nu\mathbf{1}_n,\mathbf{I}_n)}, \quad (۸)\end{aligned}$$

به دست می‌آید که در آن $\boldsymbol{h}_i^\top=(1,h_{i1},\cdots,h_{ip})$، i امین سطر ماتریس متغیرهای کمکی $\boldsymbol{H}$ و $\boldsymbol{\eta}=(\boldsymbol{\beta}^\top,\sigma,\varphi,\nu,\gamma)^\top$ بردار پارامترهای مدل است. فرض کنید $\boldsymbol{Y}^\top=(y_1,\cdots,y_k)$ بردار متغیرهای پاسخ فضایی در موقعیت‌های $\{s_1,\cdots,s_k\}$، $k\le n$ باشد. فرض کنید متغیرهای پاسخ دارای تابع چگالی $f(\boldsymbol{y}|\boldsymbol{x})$، به‌طور شرطی مستقل و از یک خانواده‌ی توزیع نمایی باشد. بنابراین در موقعیت s_i، $i=1,\ldots,k$ می‌توان نوشت

$$f(y_i|x_i)=\exp\{y_ix_i-b(x_i)+c(y_i)\}, \quad (۹)$$

که در آن $b(x_i)$ تابع کومولانت و مطابق مدل‌های خطی تعمیم‌یافته $\mathrm{E}(Y_i|x_i)=g^{-1}(x_i)$ باشد که در آن $g(\cdot)$ یک تابع پیوند معلوم است.

۱.۴ برآورد پارامترهای مدل

تابع درستنمایی مدل به‌صورت $L(\boldsymbol{\eta}|\boldsymbol{y})=\int\{\prod_{i=1}^k f(y_i|x_i)\}f(\boldsymbol{x}|\boldsymbol{\eta})d\boldsymbol{x}$ می‌شود. اکنون از رابطه‌ی (۸) و (۹) تابع درستنمایی را می‌توان به‌صورت

$$\begin{aligned} L(\boldsymbol{\eta}\mid\boldsymbol{y}) = & \int \exp\{\sum_{i=1}^k (y_ix_i-b(x_i)+c(y_i)) \\ & -\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{H\beta})^\top\boldsymbol{C}_\varphi{}^{-1}(\boldsymbol{x}-\boldsymbol{H\beta})\} \\ & \times\frac{\Phi_n(\frac{\gamma}{\sigma}\boldsymbol{C}_\varphi{}^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{H\beta}),\nu\mathbf{1}_n,(1-\gamma^2)\mathbf{I}_n)}{\Phi_n(\mathbf{0};\nu\mathbf{1}_n,\mathbf{I}_n)}dx,\end{aligned}$$

نوشت که شکل بسته‌ای ندارد و برای به دست آوردن برآوردهای بیشینه درستنمایی نمی‌توان از روش‌های معمول استفاده کرد و نیاز به استفاده از الگوریتم‌های تکرارشونده است. برای افزایش سرعت محاسبات می‌توان از تابع درستنمایی دوبه‌دویی به‌صورت

$$\begin{aligned} PL(\boldsymbol{\eta}|\boldsymbol{y}) &= \prod_{(\ell,\ell')\in\aleph} L(\boldsymbol{\eta}|y_\ell,y_{\ell'}) \\ &= \prod_{(\ell,\ell')\in\aleph}\int\int f(y_\ell|x_\ell)f(y_{\ell'}|x_{\ell'}) \\ &\times\ f(x_\ell,x_{\ell'}|\boldsymbol{\eta})dx_\ell dx_{\ell'}\end{aligned}$$

استفاده کرد که در آن $\aleph$ یک زیرمجموعه از همسایگی‌هاست. با به کار بردن این تابع درستنمایی مرکب و یک الگوریتم تکرارشونده مثل الگوریتم بیشینه‌سازی امید ریاضی می‌توان برآورد بیشینه درستنمایی مدل را به دست آورد. چون در الگوریتم بیشینه‌سازی امید ریاضی نیاز به توزیع شرطی $(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\eta})$ است. [۴] نشان دادند که در یک حالت خاص این توزیع شرطی به‌صورت تقریبی دارای توزیع چوله نرمال بسته است. ابتدا صورت قضیه اثبات‌شده توسط [۴] بیان و سپس به مدل آمیخته خطی تعمیم‌یافته فضایی تعمیم‌یافته برای میدان تصادفی چوله گاوسی مانا تعمیم داده می‌شود.

قضیهٔ ۱.۴. *فرض کنید متغیرهای پنهان فضایی در یک مدل آمیخته خطی تعمیم‌یافته فضایی دارای توزیع چوله نرمال بسته*

$$(\boldsymbol{X}|\boldsymbol{\eta})\sim CSN_{n,1}(\boldsymbol{H\beta},\boldsymbol{\Sigma}_\theta,\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}_\theta^{-\frac{1}{2}},0,1)$$

باشد و $\boldsymbol{x}$ به‌صورت $\boldsymbol{X}=(\boldsymbol{X}^{obs\top},\boldsymbol{X}^{pred\top})^\top$ افراز شده باشد. همچنین فرض کنید متغیرهای پاسخ گسسته فضایی متعلق به خانواده‌ی نمایی

$$f(y_i|x_i)=\exp\{y_ix_i-b(x_i)\},\ i=1,\cdots k$$

باشند. با در نظرگرفتن $A=[I_{k\times k}|\mathbf{0}_{k\times n-k}]$، $X^{obs}=AX$ و با خطی سازی قسمت درستنمایی $f(y|x)f(x|\eta)$ حول یک مقدار ثابت x داریم

$$X|Y,\eta \approx CSN_{n,1}(\mu_{x|y,\eta},\Sigma_{x|y,\eta},D_{x|y,\eta},\nu_{x|y,\eta},1), \quad (10)$$

که در آن

$$\mu_{x|y,\eta}=H\beta+\Sigma_\theta A^\top(A\Sigma_\theta A^\top+P)^{-1}(z(y,x^{obs})-AH\beta),$$

و $z_i(y_i,x_i)=[y_i-b'(x_i)+x_ib''(x_i)]/b''(x_i)$، $i=1,\ldots,k$ خطی سازی قسمت درستنمایی در مقدار ثابت از x است.
ماتریس $R=A\Sigma_\theta A^\top+P$ و $P=P(x)$ یک ماتریس قطری با عناصر روی قطر $P(i,i)=1/b''(x_i)$, می‌باشد. به‌علاوه

$$\begin{aligned}\Sigma_{x|y,\eta} &= \Sigma_\theta-\Sigma_\theta A^\top(A\Sigma_\theta A^\top+P)^{-1}A\Sigma_\theta,\\ D_{x|y,\eta} &= \lambda^\top\Sigma_\theta^{-\frac{1}{2}},\\ \nu_{x|y,\eta} &= \lambda^\top\Sigma_\theta^{-\frac{1}{2}}(H\beta-\mu_{x|y,\eta}).\end{aligned}$$

در ادامه قضیه‌ای را بیان می‌کنیم که رابطه خطی سازی درستنمایی مدل‌های آمیخته خطی تعمیم‌یافته فضایی تعمیم‌یافته برای میدان تصادفی چوله گاوسی مانا تعمیم بدهد.

قضیهٔ ۲.۴. *فرض کنید متغیرهای پنهان فضایی در مدل آمیخته خطی تعمیم‌یافته فضایی تعمیم‌یافته دارای توزیع چوله نرمال بسته*

$$CSN_{n,n}(H\beta,\sigma^2C_\varphi,\frac{\gamma}{\sigma}C_\varphi^{-\frac{1}{2}},\nu\mathbf{1_n},(1-\gamma^2)\mathbf{I_n}),$$

باشد و X به‌صورت $X=(X^{obs\top},X^{pred\top})^\top$ افراز شده باشد. همچنین فرض کنید متغیرهای پاسخ گسسته فضایی متعلق به خانواده نمایی

$$f(y_i|x_i)=\exp\{y_ix_i-b(x_i)\},\ i=1,\cdots k$$

باشند. آنگاه با در نظرگرفتن $A=[I_{k\times k}|\mathbf{0}_{k\times n-k}]$، $X^{obs}=AX$ و با خطی سازی قسمت درستنمایی $f(y|x)f(x|\eta)$ حول یک مقدار ثابت x، توزیع $X|y,\eta$ به‌طور تقریبی چوله نرمال بسته با پارامترهای

$$CSN_{n,n}(\mu_{x|y,\eta},\Sigma_{x|y,\eta},\frac{\gamma}{\sigma}C_\varphi^{-\frac{1}{2}},\nu_{x|y,\eta},(1-\gamma^2)\mathbf{I_n}) \quad (11)$$

است که در آن

$$\mu_{x|y,\eta}=H\beta+C_\varphi R(z(y,x^{obs})-AH\beta),$$

و $z_i(y_i,x_i)=[y_i-b'(x_i)+x_ib''(x_i)]/b''(x_i)$، $i=1,\ldots,k$ خطی سازی قسمت درستنمایی در مقدار ثابت از x است.

ماتریس $R=A^\top(AC_\varphi A^\top+\frac{1}{\sigma^2}P)^{-1}$ و $P=P(x)$ یک ماتریس قطری با عناصر روی قطر $P(i,i)=1/b''(x_i)$, می‌باشد. به‌علاوه

$$\begin{aligned}\Sigma_{x|y,\eta} &= \sigma^2C_\varphi(\mathbf{I_n}-RAC_\varphi),\\ \nu_{x|y,\eta} &= \nu\mathbf{1_n}-\sigma\gamma C_\varphi^{\frac{1}{2}}R(\mathbf{z}(y,x^{\mathbf{obs}})-AH\beta).\end{aligned}$$

اثبات. فرض کنید

$$\begin{aligned}W &\sim N_n(H\beta,\sigma^2C_\varphi),\\ E &\sim N_n(\mathbf{0_n},(1-\gamma^2)\mathbf{I_n}),\\ X_0 &= -\nu+\frac{\gamma}{\sigma}C_\varphi^{-\frac{1}{2}}(W-H\beta)+E,\end{aligned}$$

در این صورت

$$\begin{pmatrix} W \\ X_0 \end{pmatrix}\sim N_{2n}\left(\begin{pmatrix} H\beta \\ -\nu \end{pmatrix},\begin{pmatrix} \sigma^2C_\varphi & \sigma\gamma C_\varphi^{\frac{1}{2}} \\ \sigma\gamma C_\varphi^{\frac{1}{2}} & \mathbf{I_n}\end{pmatrix}\right),$$

است، آنگاه توزیع شرطی $(W|X_0>\mathbf{0})$ طبق [۳] به‌صورت

$$CSN_{n,n}(H\beta,\sigma^2C_\varphi,\frac{\gamma}{\sigma}C_\varphi^{-\frac{1}{2}},\nu\mathbf{1_n},(1-\gamma^2)\mathbf{I_n}),$$

به دست می‌آید که همان توزیع مفروض $X|\eta$ برای متغیرهای پنهان در مدل تعریف‌شده این مقاله است. حال فرض کنید $Y=AW+\epsilon$، $Y\in R^m$، که در آن A یک ماتریس $m\times n$ و $\epsilon\sim N(\mathbf{0},P)$ است. قرار دهید $Z=(Y^\top,W^\top)^\top$ آنگاه توزیع توأم Z و X_0 به‌صورت زیر حاصل می‌شود:

$$\begin{pmatrix} Z \\ X_0 \end{pmatrix}\sim N_{m+2n}\left(\begin{pmatrix} \mu_z \\ -\nu \end{pmatrix},\begin{pmatrix} \Sigma_z & \sigma\gamma A^{*\top}C_\varphi^{\frac{1}{2}} \\ \sigma\gamma C_\varphi^{\frac{1}{2}}A^* & \mathbf{I_n}\end{pmatrix}\right),$$

که در آن

$$\mu_z=\begin{pmatrix} AH\beta \\ H\beta \end{pmatrix},\Sigma_z=\begin{pmatrix} \sigma^2AC_\varphi A^\top+P & \sigma^2AC_\varphi \\ \sigma^2C_\varphi A^\top & \sigma^2C_\varphi\end{pmatrix},$$

و $A^*=(A'\ \ \mathbf{I_n})$ یک ماتریس $n\times(m+n)$ بعدی است.
اکنون توزیع شرطی به‌صورت

$$Z|X_0\sim CSN_{m+n,n}(\mu_z,\Sigma_z,\frac{\gamma}{\sigma}C_\varphi^{-\frac{1}{2}}A^*\Sigma_z^{-1},\nu\mathbf{1_n},\mathbf{\Delta_z})$$

به دست می‌آید که در آن $\Delta_z=\mathbf{I_n}-\frac{\gamma^2}{\sigma^2}C_\varphi^{-\frac{1}{2}}\mathbf{A}^*\mathbf{\Sigma_z}^{-1}C_\varphi^{-\frac{1}{2}}$ است و چون داریم

$$\Sigma_z^{-1}=\begin{pmatrix} P^{-1} & -P^{-1}A \\ -A^\top P^{-1} & \frac{1}{\sigma^2}C_\varphi^{-1}+A^\top P^{-1}A\end{pmatrix},$$

پس طبق خواص توزیع چوله نرمال بسته، حاشیه‌ای $(W|Y,X_0>\mathbf{0})$

$$CSN(H\beta+C_\varphi R(y-AH\beta),\sigma^2C_\varphi-\sigma^2C_\varphi RA\sigma^2C_\varphi,$$

$$\frac{\gamma}{\sigma}\boldsymbol{C}_{\varphi}^{-\frac{1}{2}}, \nu\mathbf{1_n} - \sigma\gamma\boldsymbol{C}_{\varphi}^{-\frac{1}{2}}\boldsymbol{R}(\boldsymbol{y}-\boldsymbol{AH\beta}), (\mathbf{1}-\gamma^2)\mathbf{I_n}),$$

حاصل می‌شود. پس توزیع $(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{\eta})$ که معادل توزیع $(\boldsymbol{W}|\boldsymbol{Y},\boldsymbol{X}_0>0)$ است، یک توزیع چوله نرمال بسته به‌صورت رابطه (۱۱) است. □

اکنون با استفاده از الگوریتم بیشینه‌سازی امید ریاضی و قضیه (۲.۴) یک الگوریتم به شرح زیر برای برآورد پارامترهای مدل پیشنهادی معرفی می‌شود.

گام ۱- در مرحله اول مقادیر اولیه $\boldsymbol{\eta}^{(0)}$ طوری انتخاب می‌شوند که $PL(\boldsymbol{\eta}^{(0)}|\boldsymbol{y})>0$ باشد و m را صفر قرار می‌دهیم.

گام ۲- در مرحله دوم یک مقدار اولیه برای $\boldsymbol{x}^{(0)}$ را به‌صورت مد $f(\boldsymbol{x}_{\ell\ell'}|\boldsymbol{\eta})$ فرض و $d=0$ در نظر گرفته می‌شود.

گام ۳- با استفاده از قضیه ۲.۴ توزیع شرطی $\hat{f}(\boldsymbol{x}_{\ell\ell'}|\boldsymbol{y}_{\ell\ell'},\boldsymbol{\eta})$ محاسبه و سپس $\boldsymbol{x}^{(d)}$ را مد توزیع $\hat{f}(\boldsymbol{x}_{\ell\ell'}|\boldsymbol{y}_{\ell\ell'},\boldsymbol{\eta})$ در نظر بگیرید و d را $d+1$ قرار دهید و تا رسیدن به همگرایی مرحله دوم را تکرار و شکل توزیع شرطی مشخص می‌شود.

گام ۴- اکنون از الگوریتم بیشینه‌سازی امید ریاضی عبارت

$$\begin{aligned} Q(\boldsymbol{\eta}|\boldsymbol{\eta}^{(m)}) &= \sum_{(\ell,\ell')\in\aleph}\int\int \log\Big\{f(x_\ell,x_{\ell'},y_\ell,y_{\ell'}|\boldsymbol{\eta})\Big\} \\ &\times\hat{f}\Big(x_\ell,x_{\ell'}|y_\ell,y_{\ell'},\boldsymbol{\eta}^{(m)}\Big)dx_\ell dx_{\ell'}, \quad (12) \end{aligned}$$

محاسبه و $\boldsymbol{\eta}^{(m+1)}$ را طوری انتخاب می‌کنیم که

$$\boldsymbol{\eta}^{(m+1)} = \text{argmax}_{\boldsymbol{\eta}} Q(\boldsymbol{\eta}|\boldsymbol{\eta}^{(m)}).$$

الگوریتم را تا زمان رسیدن به همگرایی ادامه می‌دهیم و برآورد بیشینه درستنمایی پارامترها مشخص می‌شود.

۵ مطالعه شبیه‌سازی

روی نقشه‌ی ایران و به‌صورت یک مشبکه‌ی نامنظم ۱۰۰ موقعیت تولیدشده است. با در نظر گرفتن ساختار همسانگرد نمایی برای $\boldsymbol{C}_\varphi$ و

$$\beta_0=-2,\ \beta_1=1,\ \sigma^2=1,\ \varphi=6,\ \gamma=0.95,$$

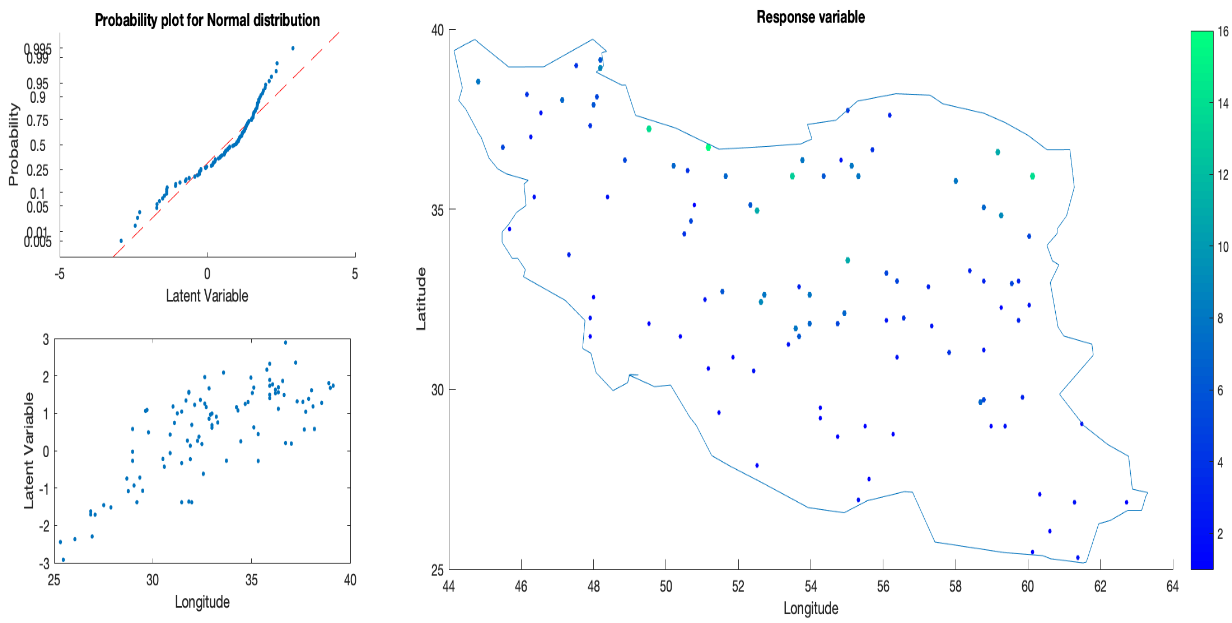
متغیرهای پنهان از توزیع چوله نرمال به‌صورت

$$(\boldsymbol{x}|\boldsymbol{\eta}) \sim CSN_{n,n}(\beta_0+\beta_1\boldsymbol{h}, \sigma^2\boldsymbol{C}_\varphi, \frac{\gamma}{\sigma}\boldsymbol{C}_\varphi^{-\frac{1}{2}}, \nu\mathbf{1_n}, (1-\gamma^2)\mathbf{I_n})$$

استخراج شد که در آن $\boldsymbol{h}$ بردار مقادیر استانداردشده عرض جغرافیایی است که به‌عنوان متغیر کمکی وارد مدل می‌شود. متغیر پاسخ به‌شرط متغیرهای پنهان، $y_j, j=1,\cdots,100$ از توزیع پواسون به‌صورت

$$y_j \sim Poiss(p_j),\ p_j=\exp(x_j)$$

تولید و فرایند شبیه‌سازی ۱۰۰۰ بار تکرار شد. یک نمونه از موقعیت‌ها و داده‌های تولیدشده در شکل ۱ رسم شده است. برای ۱۰۰۰ مجموعه داده‌ی تولیدشده و مدل چوله گاوسی مانا و مدل گاوسی، رهیافت درستنمایی معرفی‌شده در مقاله اجرا شد و خلاصه نتایج به شرح جدول ۱ ارائه‌شده است. نتایج نشان می‌دهد که با به کار بردن میدان تصادفی مانا و رهیافت درستنمایی دودویی معرفی‌شده برای همه پارامترها متوسط میانگین خطا برآورد همه‌ی پارامترها برای مدل چوله گاوسی و همچنین مدل گاوسی به‌خوبی عمل نموده است.



شکل ۱. تحققی از مدل پواسون فضایی شبیه‌سازی‌شده: راست) داده‌های پاسخ پواسونی، چپ-بالا) چندک-چندک متغیرهای پنهان، چپ-پایین) نمودار پراکنش متغیرهای پنهان در مقابل عرض جغرافیایی.

جدول ۱. نتایج شبیه‌سازی بر اساس ۱۰۰ × ۱۰۰۰ مجموعه داده‌ی تولیدشده.

گاوسی			چوله گاوسی				
Ave. MSE	Ave. Std.	Ave. Es.	Ave. MSE	Ave. Std.	Ave. Es.	R. val.	Par.
۰٫۸۵۹۹	۰٫۱۳۹۸	−۱٫۶۷۱۹	۰٫۶۵۶۱	۰٫۱۰۳۲	−۱٫۹۴۴۴	−۲	β_0
۰٫۴۳۱۱	۰٫۱۱۰۱	۱٫۳۳۹۰	۰٫۳۲۰۹	۰٫۰۸۸۹	۱٫۰۷۹	۱	β_1
۰٫۱۰۰۳	۰٫۰۸۰۹	۱٫۴۳۰۰	۰٫۴۶۰۰	۰٫۰۵۱۸	۱٫۲۰۰۸	۱	σ^2
۲٫۱۵۰۲	۰٫۶۸۴۱	۷٫۱۵۰۱	۱٫۶۶۹۱	۰٫۳۱۱۱	۶٫۲۰۲۹	۶	φ
−	−	−	۰٫۰۵۰۰	۰٫۰۱۶۰	۰٫۸۳۱۱	۰٫۹۵	γ

بحث و نتیجه‌گیری

در این مقاله مدل‌های آمیخته خطی تعمیم‌یافته فضایی بر اساس یک میدان تصادفی چوله گاوسی مانا مدل‌بندی شد که این میدان تصادفی دارای شرط سازگاری حاشیه‌ای هست. در یک قضیه اثبات شد که با مدل‌بندی متغیرهای پنهان با میدان تصادفی چوله گاوسی مانا توزیع شرطی متغیرهای پنهان را می‌توان با یک توزیع چوله نرمال بسته تقریب زد. با این توزیع شرطی تقریبی به‌دست‌آمده و الگوریتم بیشینه‌سازی امید ریاضی تقریبی الگوریتمی برای به دست آوردن برآورد بیشینه درستنمایی پارامترهای مدل بر اساس یک رهیافت درستنمایی مرکب معرفی شد. نتایج شبیه‌سازی نشان می‌دهد که مدل معرفی‌شده بر اساس میدان تصادفی چوله گاوسی مانا و رهیافت درستنمایی به دست آوردن برآوردهای مدل، به‌خوبی عمل نموده است.

مراجع

[۱] کریمی، ا. و حسینی، ف. (۱۴۰۰)، معرفی یک میدان تصادفی مانای چوله گاوسی، *مجله علوم آماری ایران*، ۱۵، ۵۴۹-۵۶۶.

[2] Allard, D., and Naveau, P. (2007), A New Spatial Skew-Normal Random Field Model, *Communications in Statistics—Theory and Methods*, **36**, 1821-1834.

[3] Gonzalez-Farias, G., Dominguez-Molina, A., and Gupta, A. K. (2004), The Closed Skew Normal Distribution. In: *Genton M. G., ed. Skew-elliptical distributions and their applications: A journey beyond normality*. Boca Raton, FL: Chapman and Hall CRC, 25-42.

[4] Hosseini, F., Eidsvik, J., and Mohammadzadeh, M., (2011). Approximate Bayesian inference in Spatial GLMM with Skew Normal Latent Variables, *Computational Statistics and Data Analysis*, **55**1791-1806.

[5] Hosseini, F., and Mohammadzadeh, M., (2012). Bayesian prediction for spatial GLMM's with Closed Skew Normal latent variables, *Australian & New Zealand Journal of Statistics*, **54**, 43-62.

[6] Hosseini, F. (2016). A new algorithm for estimating the parameters of the spatial generalized linear mixed models, *Environmental and Ecological Statistics*, **23**, 205-217.

[7] Hosseini, F., and Karimi, O. (2019). Approximate composite marginal likelihood inference in spatial generalized linear mixed models, *Journal of Applied Statistics* **46(3)**, 542-558.

[8] Hosseini, F., and Karimi, O. (2020). Approximate likelihood Inference in Spatial Generalized Linear Mixed Models with Closed Skew Normal latent Variables, *Communication in Statistics- Simulation and Computation* **49**, 121-134.

[9] Hosseini, F., and Karimi, O., (2021). Approximate pairwise likelihood inference in SGLM models with skew normal latent variables, *Journal of Computational and Applied Mathematics*, **398**, 113692.

[10] Kim, H.M., and Mallick, B.K., (2004). A Bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning and Inference*, **120**, 85–101.

[11] Karimi, O., Omre, H., and Mohammadzadeh, M., (2010). Bayesian Closed-skew Gaussian Inversion of Seismic AVO Data for Elastic Material Properties, *Geophysics,* **75**, R1-R11.

[12] Karimi, O., and Mohammadzadeh, M., (2011). Bayesian Spatial Prediction for Discrete Closed Skew Gaussian Random Field, *Mathematical Geosciences* **43**, 565–582.

[13] Karimi, O., and Mohammadzadeh, M., (2012). Bayesian spatial regression models with closed skew normal correlated errors and missing observations, *Statistical Papers* **53(1)**, 205-218.

[14] Rimstad, K., and Omre, H. (2014). Skew-Gaussian Random Fields, *Spatial Statistics* **10**, 43-62.

[15] Varin, C., Reid, N., and Firth, D., (2011) . An overview of composite likelihood methods, *Statistica Sinica*, **5**, 42, 2011.

On Estimation for Spatial Generalized Linear Mixed Models using a Stationary Skew Gaussian Random Field

Fatemeh Hosseini[1] and Omid Karimi [2]

Abstract:

Spatial generalized linear mixed model is commonly used to model Non-Gaussian data and the spatial correlation of the data is modelled by latent variables. In this paper, latent variables are modeled using a stationary skew Gaussian random field and a new algorithm based on composite marginal likelihood is presented. The performance of this stationary random field in the model and the proposed algorithm is implemented in a simulation example.

Keywords: Spatial generalized linear mixed model, stationary skew Gaussian random field, composite marginal likelihood.

[1]**Semnan University, Department of Statistics**
[2]**Semnan University, Department of Statistics**